Thesis/
Reports
Nowell,
C.

Encounter Sampling in the Bridger-Teton National Forest

COOP AGREEMENT WEBER STATE UNIV

FINAL REPORT FOR RESEARCH AGREEMENT
#INT-95057-RJVA
"Monitoring Wilderness Visitors Using Encounter Sampling"
WEBER STATE UNIVERSITY
FS Contact: Dr. David L. Turner
CoOp Contact: Dr. Cliff Nowell

Encounter Sampling in the Bridger-Teton National Forest

Final Report

November 1998

Cliff Nowell
Department of Economics
Weber State University
Ogden, UT. 84403

David L. Turner
USDA Forest Service Research
Rocky Mountain Research Station
860 North 1200 East
Logan, Utah 84321

Abstract

Obtaining information on the visitors who use the National Forests is a difficult sampling problem. Researchers often rely on trail register data to estimate usage and to estimate visitor characteristics. In this study we discuss the bias contained in trail registration data in the Bridger-Teton National Forest in western Wyoming and present a way to remove this bias. We then evaluate an alternative sampling scheme whereby data is gathered by wilderness rangers as the travel through the back country. Although it is well known that data obtained through a sampling plan such as this is length-biased, we show that even after correcting for this bias it is unlikely that this data can be used to estimate visitor usage.

I. Background

Wilderness ranger reports are an often used method of gathering data in the Bridger-Teton National Forest in western Wyoming. The purpose of the wilderness ranger reports are varied; however, both visitor characteristics and usage estimates are of interest to the Bridger-Teton National Forest. Our goal in this study is to make use of wilderness ranger reports in estimating total back country use. To accomplish this task we do two things. First, we determine how back country ranger counts can be used to adjust the data obtained from the trail registers. Second, we make use of the relationship between a recreationist's length of stay in the back country and the probability of the recreationist being observed by the wilderness ranger, to adjust the data observed by the wilderness ranger.

To gather data, the typical wilderness ranger undertakes multiple day, back country trips which originate from a predetermined point and follow a predetermined route. These trips have

multiple objectives, one of which is to gather data from visitors. The wilderness ranger records data whenever individuals are encountered. Such an approach can cause back country users to be sampled with a probability proportional to their length of stay in the wilderness (Cox, 1969). This will cause the data collected on length of stay to be unreliable. If the other variables of interest such as group size and group destination are correlated with the visitors length of stay, their estimates will also be biased.

In this study we attempted to determine whether the sample collected by a typical wilderness ranger is likely to be length biased. We also examine if visitor characteristics are correlated with length of stay. Based on data collected in the Bridger-Teton National Forest in 1995, we compare the length biased sample collected through encounter sampling with the sample collected from the trailhead registration boxes.

The study area extends approximately 80 miles along the Continental Divide on the west slope of the Wind River Mountain Range. The Wind River range contains 48 summits higher than 12,500 feet including Gannet Peak, the highest peak in Wyoming. The Bridger-Teton National Forest is generously endowed with thousands of lakes and ponds and contains the headwaters of the Green River. The area recorded a quarter million recreation visitor days on the back country trails in 1994.

II. Estimating Back country Users based on Trail Register Data.

Currently, the Bridger-Teton National Forest estimates total usage by using data from trail registers and inflating these numbers to account for the fact that a significant percent of users do not sign the trail registers. One purpose of our study is to calculate a new factor that can be used

to adjust the trail registers. We will address this issue before we discuss the data collected through wilderness ranger reports. We now turn to this task.

During 1996 and 1997, surveyors asked back country users if they had registered at the trailhead. Of the 546 people surveyed, 81.32 percent claimed they had registered. As a preliminary analysis, the Classification and Regression Tree procedure developed by Brieman et al. (1984) as implemented by the CART (1998) program, version 3.5 was employed. Using the variables Days (or length of the trip), Size (of the group) and Foot (1 if on foot, 0 if horseback), the CART produced the following classification tree to predict whether encountered users had registered or not.

CART uses the entire data set to develop a model and then validates the fit by repeatedly deleting 10% of the sample and refitting the model. For this data set, the final model is displayed in the following figure. The first node (at the top of the figure) says that 102 of 546 in the original sample said they were not registered. The sample is then divided into two groups, the right group containing those staying more than 2.5 days (i.e. 3 or more days since days were measured to whole days only) and the left staying less than 2.5 (i.e. 0, 1 or 2 days). Cases falling into the box on the right are classified as a "1", indicating they said they registered. This is a terminal node and of the 469 observations classified into this node, only 62 said they are not registered, for an error rate (for this node) of 62/469 or 13.2%.

The remaining analysis applies to 77 of the original 546 subjects initially classified as non-registrants. Of those, the Foot variable was found to be useful in further separation of the groups. Of the 10 horseback users (Foot = 0) in this group of 77, most 8 failed to register, so the model would predict horseback riders staying less than 2.5 days would likely not register. The

2/10 under this terminal node indicates that 20% of the horseback riders staying less than 2.5 days said they registered.

Finally, of the hikers (Foot = 1), those staying more than 1.5 days (or 2 days since this branch is already constrained to be less than 2.5 days), 18 of 20 registered. If we predict that this group registered, then we would make an error 12% (6/20) of the time. For day hikers, 26 of 47 failed to register, making an error rate for this group of 21/47 or 44.7% if we predicted the group would not register.

The overall misclassification rate is 62 + 2 + 21 + 6 or 91 out of 546 or 16.7%. Since this error rate is computed using all the data, it is often a little too optimistic. To adjust for this, CART then omits 10% of the original data and refits the model. After repeating this 12 times, the crossvalidation misclassification error rate is 101 of 546 or 18.5%.

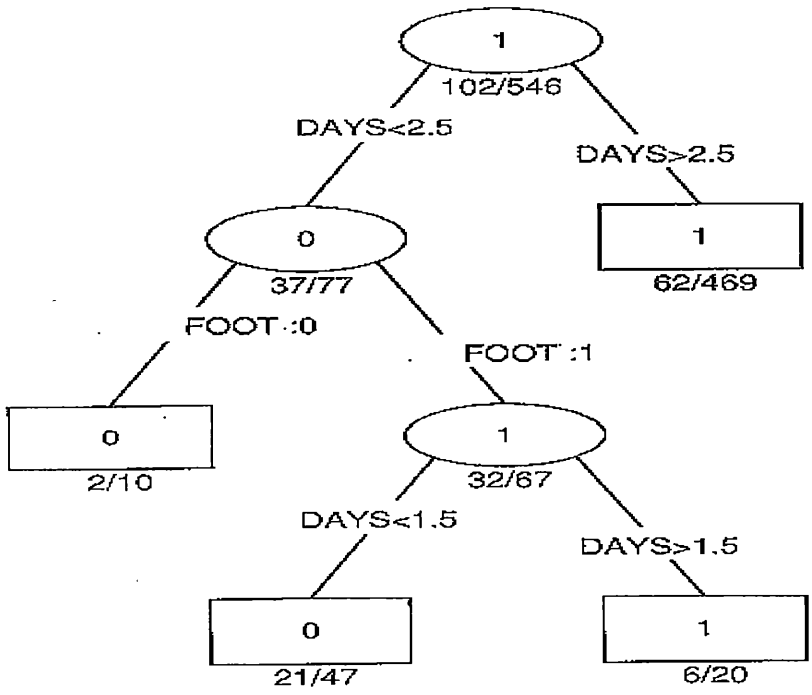


Figure 2. CART model.

It is clear that different types of visitors have different probabilities of registering. We explore this further immediately below.

To further explain why some types of groups were more likely to register than others we estimate a logistic regression model to predict the probability of any individual registering at the trailhead. We used the data collected at the trail registers. The dependent variable is a binomial variable which equals one if the respondent indicated they had registered at the trailhead and zero otherwise. The explanatory variables are whether the respondent was on foot or had stock (FOOT=1 without stock and FOOT = 0 with stock), the number of days the respondent planned to stay in the back country (DAYS) and the size of the responding group (SIZE). We estimate the probability that an individual registers at the trailhead (P_i) by using the logistic distribution so that $P_i = [1/1 + e^{-(\beta 0 + \beta 1 Foot + \beta 2 DAYS + \beta 3 Size)}]$.

Results of the logistic analysis are presented in Table One. The data used to estimate the equation are the 546 observations collected during the last two years of the study.

Table One: Logistic Analysis

Variable	Estimated Coefficient	p-value	Mean
Constant	-0.692	0.09	1.00
Foot	1.38	0.00002	0.899
Days	0.15	0.002	5.31
Size	0.065	0.16	3.60
n=546			

Overall, the logistic analysis predicts registration well as indicated by the global chi-

square test statistic of 17.52 which is significant at less than the .01 level. The average predicted probability of registering is calculated using $P_i = [1/1 + e^{-(\beta 0 + \beta 1 Foot + \beta 2 DAYS + \beta 3 Size)}]$. Using the mean values for all explanatory variables results in $P_i = .8291$. Regression analysis with a probit specification does not significantly alter the results presented above.

Using the estimates from Table One and data collected from the trail register we are able to calculate the predicted number of users in the Bridger-Teton National Forest. Because the data collected at the trail register is from a group of back country users with different characteristics than the characteristics of the surveyed group we used a sample from the trail registers of 500 users during 1996 and 1997 to estimate the total number of recreationists in the forest. Using the the mean value for all data from the trail registers (FOOT= .986, DAYS = 2.82, SIZE = 2.92), we calculate Pi = 1/[1 + exp(-(-.692 + 1.38FOOT + .15DAYS +.065SIZE))]. Based on these calculations we find Pi = .7827. Thus we conclude that trail registers should be adjusted upwards by 27.86 percent. Thus to accurately estimate usage, the number of visitors on the trail registered should be multiplied by 1/0.7827 = 1.2786.

III. Length-Biased Sampling and Visitor Usage

A sample is said to be length-biased if the probability of observing an individual at a particular site is proportional to the individual's length of stay at the site. Clearly the chances of a wilderness ranger observing a recreationist in the back country are proportional to the amount of time the individual spends in the back country. Thus, visitor estimates based on wilderness ranger encounters are likely to be length biased. In this case-- and if an agency is interested in collecting visitor statistics such as: group size, place of origin, length of stay, or any other

demographic or usage estimates-- the collected data will be biased.

To outline the method used to adjust length biased data we follow Nowell, Evans, and McDonald (1988). Let X_i denote the length of visit to the site during the time interval {0, L) by the i^{th} visitor, i = 1,2,...,N, where the total number of visitors, N, is known. Also, let parameters associated with a secondary attribute, such as group size or group destination, be denoted by the random variable *Y*. Let $\{f_x(o), f_y(o)\}$ denote the probability density functions of the natural (unbiased) distributions of *X* and *Y*. Similarly, let $\{g_x(o), g_y(0)\}$ represent the corresponding probability density functions of the observed (biased) data resulting from the wilderness ranger report. Expectations with respect to the natural distributions will be denoted by $E_f(X) = u_x$, $E_f(Y) = u_y$. Expectations with respect to the observed distributions are denoted by $E_g(o)$.

If the researcher knows the total length of stay in the back country for each individual encountered, appropriate parameter estimation was shown by Cox(1969). The fundamental results are:

$$g_x(x) = xf_x(x)/u_x \qquad 0 \le x \le \infty \qquad (1)$$

$$g_y(y) = E_f(x|y)f_y(y)/u_x. \qquad -\infty < y < \infty \qquad (2)$$

Equation (1) states that the observed, biased distribution of *X* equals the natural, unobserved distribution of *X* weighted by the ratio of the length of visit of an individual to the unbiased mean length of visit. Rearranging this equation gives:

$$f_x(x) = g_x(x)(u_x/x) \qquad (3)$$

From (3) it is easy to see that the observed data, $g_x(x)$ should be assigned weights inversely proportional to the length of stay. The relative importance of sampled individuals who spend large amounts of time is discounted, because they are likely to be over represented in the

encounter sample, and the importance of individuals who spend a small amount in the Bridger-Teton National Forest is increased because they are likely to be under represented in the encounter sample. Equation (2) has a similar interpretation to Equation (1). The observed distribution is equal to the true, unbiased distribution weighted by a factor that takes into account the individual's length of stay. In order to apply the above formulae it is necessary to have an estimator of u_x. Estimators for u_x and u_y are given by:

$$\hat{(1/u_x)} = 1/n\{\sum_{1}^{n} 1/x_i\} \quad (4)$$

$$\hat{u}_y = \sum_{1}^{n}(y_i/x_i)/\sum_{1}^{n}(1/x_i) \quad (5)$$

where n is the number of individuals encountered in the length biased sample. Equations (5) and (6) give the necessary information to calculate the unbiased estimators for the mean length of stay of visitors to the Bridger-Teton National Forest and any associated parameter.

IV. Survey Data

During the summers of 1995, 1996, and 1997 researchers spent a total of 90 days in the Bridger -Teton National Forest gathering data. Data was gathered between the second week of July and the second week of September. Data was primarily gathered at three different locations: Big Sandy, Elkhart Park, and Green River Lakes. In addition, a small amount of data was gathered at the North Fork trailhead A total of 648 observations were collected. The average number of encounters was approximately seven per day. This number varied from less than one encounter per day for the time period between July 13, and July 18, 1995 to just over 15 per day between August 4, and August 11, 1997. Surveyors estimated that 60 to 70 perent of

encounters resulted in recorded data. The percentage of visitors actually surveyed was greatly effected by weather conditions and time of day. Table Two shows total usage of the Big Sandy and Green River Lakes trailhead during 1993 and 1994, the two years prior to the beginning of the study, which was obtained from the trail registers located at each trail.

Table Two: Trail Usage 1993/1994

	1993		1994	
	People	Stock	People	Stock
Big Sandy	2963	76	2739	96
Green River Lakes	2638	44	2795	70

V. Survey Methods

During the summer of 1995 two individuals spent a total of ten days in the field in the third week of July surveying at the North Fork trailhead. This trailhead was selected because it was one of very few trails where individuals had access to the back country. Even here, snow conditions were not good. Due to heavy snow conditions these individuals never got more than about ten miles from the trailhead. On the first day of the survey, although hikers were encountered by the surveyors, no data was collected. It was judged that stopping to talk with the hikers in the rain was inappropriate. It is apparent that usage estimates derived from wilderness ranger reports will be subject to the weather conditions that exist while the wilderness ranger is in the field.

One individual spent four days in the field during the second week of July at the Big

Sandy entrance. Snow conditions prohibited getting more than seven to eight miles from the trailhead. Weather conditions (rain) also limited the amount of surveys gathered. This same individual spent five days in the field during the fourth week of July at the Big Sandy entrance. Weather conditions had improved by this time and approximately eight observations were gathered per day.

One individual spent five days in the field at Big Sandy during the second week of August. This sample period yielded the greatest number of encounters per day of the year. This same individual spent five more days in the field at Green River Lakes during the third week of August.

Because of the limited data gathered during the summer of 1995, we developed some estimates of the correlation between length of stay and number of encounters. Although the sample size was small enough to inhibit detailed analysis we did observe a correlation between length of stay and size of group and length of time spent in the back country.

During 1996 the snow conditions were much improved from the prior year. In this summer we focused on the relationship between the trail register data and the encounter samples taken in the field. During this summer we began gathering data on whether visitors had signed in at the trail register.

During this summer two individuals each spent 20 days in the Bridger-Teton forest during the last week of July and the third week of August. Both individuals gathered data at the Green River Lakes trailhead and at the Elkhart Park trailhead. The average number of encounters per day during the summer was approximately eleven in the Elkhart Park area and seven in the Green River Lakes region. Surveyors estimated they interviewed 75 percent of all parties they

encountered. This was higher than the previous year, primarily because of the improved weather conditions.

During 1997 three different surveyors spent time in the forest. During this summer we focused on two additional areas of interest to the Pinedale ranger station. First we asked back country users how many other groups they had encountered during the day to focus on management objectives in different portions of the back country. Second we asked individuls if they would have been willing to pay for a back country permit for their trip. We asked if they would have been willing to pay $1, $2, $3, $4, or $5 for the permit.

During this summer two individuals spent 15 days each in the forest during the first two weeks of August, and another individual spent one week in the back country during the first week of September (including the Labor Day weekend). Individuals again surveyed Elkhart Park and Big Sandy. The average number of encounters were greatest during this summer. During the initial trip to Elkhart Park approximately 15 individuals were surveyed per day.

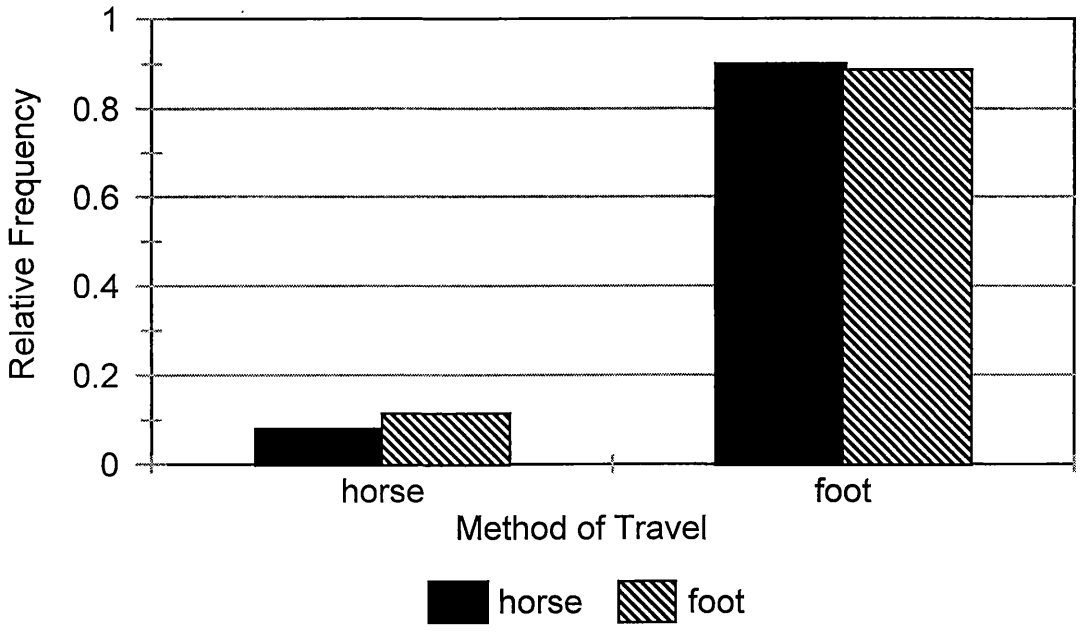
All samples were obtained in a manner designed to mimic the routine of the wilderness ranger in the Bridger-Teton National Forest. Back country users were only interviewed when it appeared to the surveyor that the interview would be non-obtrusive to the visitor. During each summer, the surveyors gathered data on the information visitors are asked to report at the trailhead registration boxes: state of origin, zip code, date of arrival, planned date of departure, and group size.

VI. Encounter Sampling

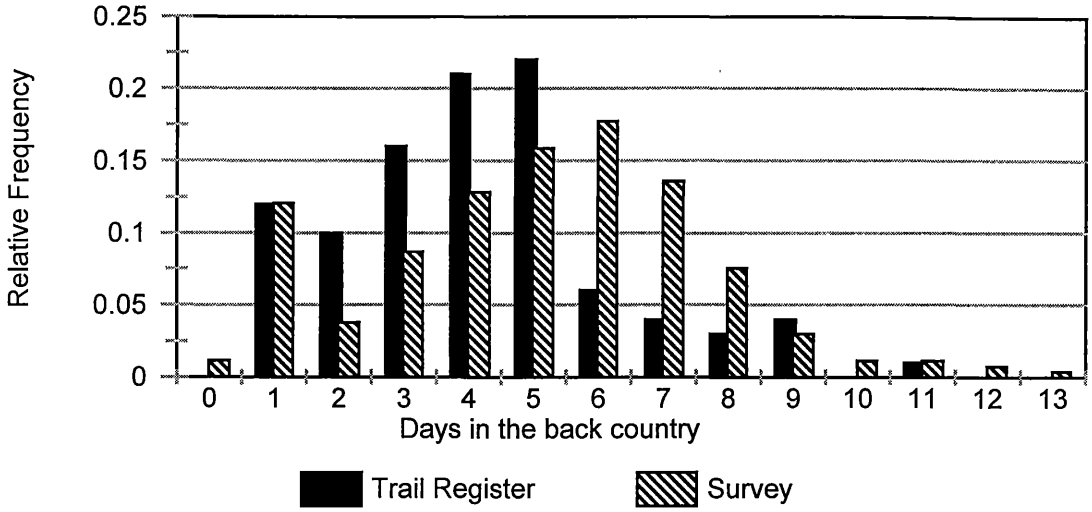
The figures below compare the distributions of trail register data and surveyed data for

three different variables: Method of travel, length of stay in the back country, and group size. The three histograms presented below illustrate the differences between data gathered at the trailhead and survey data.

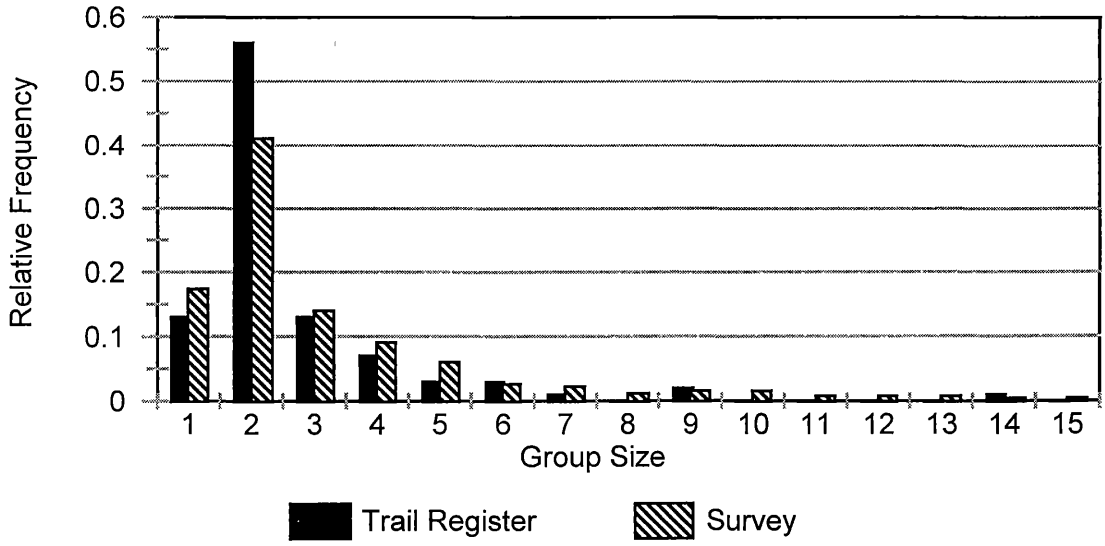
The first figure shows a substantial difference between trail register data and survey data for the method of travel. At the trailhead less than one percent of those registered claim they have stock. In the field however, about ten percent of the groups encountered have stock.



It is also clear that a large difference exists in the profiles of number of days spent in the back country when comparing data gathered in the field when compared with the data gathered at the trail register. Back country rangers are much more likely to survey individuals who spend large amounts of time in the back country.



Differences between the encounter sample and trailhead sample are not as apparent for the variable group size. Although group size is directly correlated with length of stay the differences in encounter data and trail register data is not as great.



Using the formulae presented above we present estimates of how the number of days and group size should be adjusted to compare the encounter data with the trail register data. The results are shown in Table Three.

Table Three: Adjusted Data

Variable	ρ	Biased Mean Response	Unbiased Mean Response	Trail Register Mean Response
Length of Stay	-	5.36	4.23	2.97
Group Size	.21	3.16	2.82	2.53

As shown in Table Three, data gathered by a wilderness ranger is likely to overestimate the average group size and the average length of stay of groups. Even after adjustment due to the length -biased sample these estimates are greater than the estimates obtained from the trail register.

We now turn to the problem of estimating daily visitation to the areas surveyed. The average number of encounters per day by the surveyors was slightly less than seven. This number is less than one quarter of the average number of people who register at the trailhead per day. Why is this so? Obviously because on days when the wilderness ranger is in the back country a significant distance he or she misses all of the individuals who only spend a short amount of time on the trail. The unbiased estimate of visitation would only be obtained if the wilderness ranger were to stay at the trailhead and count the number of groups as they entered the forest. That is, a relationship exists between the number of encounters per day and the distance the surveyor is from the trailhead. We now turn to the task of estimating this relationship.

We use the data collected to form a new variable which equals the number of encounters the surveyor made each day on the trail. In total we were able to obtain 364 valid observations from the data. The reduction in data size happens because of the inability to determine the number of encounters per day for some days, and a lack of data on distance from trailhead for some observations. For each observation in the data set we have a new variable equal to the number of encounters on the day the observation was taken. We posit that this number (DLYOBS) is a function of how far away from the trailhead the observation was taken (DIST). In the most simple form we use a simple regression equation to estimate $DLYOBS = \beta_0 + \beta_1 DIST$.

In this case β_0 represents the expected value of DLYOBS conditional upon DIST = 0. In other words this is the expected number of encounters per day if the surveyor were to be at the trailhead. Table Four presents the results of the regression analysis for three different functional forms.

Table Four: Regression Results

Functional Form	Estimated Coefficients	R-Square
$DLYOBS = \beta_0 + \beta_1 DIST.$	$\beta_0 = 11.76;\ \beta_1 = -.528.$	.39
$DLYOBS = \beta_0 + \beta_1 \ln(DIST).$	$\beta_0 = 11.62;\ \beta_1 = -2.14$	.36
$DLYOBS = \beta_0 + \beta_1 DIST + \beta_2 DIST^2.$	$\beta_0 = 12.47;\ \beta_1 = -.844;$ $\beta_2 = .022$	.41

n=364 in all equations
All estimated coefficients are significant at the .05 level

Table Four gives an estimate for the number of groups entering the back country per day at the trails surveyed. Two strong caveats must be attached to this number. First, because

surveyors estimated that they did not interview approximately one third to one quarter of visitors these numbers are obviously underestimates of the number of people entering the back country. Second, because the surveyors were only taking surveys for 6-10 hours in the day the numbers presented here are based on samples from only part of the time where visitors enter the back country.

VII. Willingness to Pay Estimates and Encounters with Other Groups

The table below shows the percent of individuals willing to pay for a back country permit for different levels of cost.

Table Five: Willingness to Pay

Cost of Permit	Percentage Willing to pay Cost
$1	.903
$2	.921
$3	.864
$4	.824
$5	.887

Overall, the vast majority of individuals are willing to pay for a back country permit. In addition, because increasing the price of the permit is very weakly correlated with willingness to pay, it is unlikely that price is a deciding factor in recreationists' decisions on this matter. In fact, many of the individuals who were not willing to pay for a permit indicated they thought it was in some sense unfair, and indicated philosophical reasons for their unwillingness to pay.

In addition, during the final summer of the study visitors were asked "How many other groups have you seen in the back country today." These responses were then broken down by management area and are given in Table Six below. Although very few individuals were surveyed in Management Area B, crowding is less apparent in this area. No significant differences exist in crowding between Management Areas C and D.

Table Six: Encounters with Other Groups

Management Area	Sample Size	Mean Groups Encountered
B	17	3.10
C	153	6.91
D	168	7.72

VIII. Conclusion

Wilderness rangers are not likely to be able to gather data that can be used to estimate total visitor usage in the Bridger-Teton National Forest. The reasons for this are as follows : 1) Wilderness ranger surveys are designed to be as non-invasive as possible. When visitors are actively participating in an activity such as cooking, cleaning, or making camp, wilderness rangers are most likely not going to interview the individuals. 2) Wilderness ranger reports are dependent on the weather. In inclement weather visitors are more likely to stay in their tent or to be in a hurry to get to their destination. As a result, the wilderness ranger encounters fewer

people and is not likely to inteview those encountered. 3) Because surveys are not conducted throughout the entire day, visitation estimates are only based on part of the day. Although data collected can be adjusted to eliminate the length-biased nature of the data, it is unlikely that these reports can be used as the sole source of visitation estimates.

Although we conclude wilderness ranger reports are not a good source of data for actually counting the number of people using the forest, they do provide valuable information. Particularly, these reports can be used to monitor the validity of data collected at the trail register. We have demonstrated in this paper how these reports can be used as an effective tool to adjust the trail register data.

In addition, in this report we have shown that in the Bridger -Teton National Forest, the vast majority of users would support a fee for using the back country. Imposing fees on users of $1 to $5 dollars would have very little impact on visitation.

References

Brieman, Leo, Jerome Friedman, Richard Olshen and Charles Stone. *Classification and Regression Trees*. Pacific Grove: Wadsworth, 1984.

Cox, D.R. 1969. "Some Sampling Problems in Technology," In *new Developments in Survey Sampling*, N.L. Johnson and H. Smith Eds., New York: Wiley-Interscience.

Nowell, C.R., M. Evans and L. McDonald 1988. "Length-Biased Sampling in Contingent Valuation Studies," *Land Economics*, Vol 64 No. 4, 367-371.

Steinberg, Dan, Phillip Colla and Kerry Martin. *CART–Classification and Regression Trees: Supplementary Manual for Windows.* San Diego, CA: Salford Systems, 1998.